Mon. Not. R. Astron. Soc. **000**, 1–16 (2002) Printed 3 June 2005 (MN LATEX style file v2.2)

Selfgravity and QSO disks

Jeremy Goodman
Princeton University Observatory, Princeton, NJ 08544, USA

Received ??

ABSTRACT
It is well known that the outer parts of QSO accretion disks are prone to selfgravity if heated solely by orbital dissipation. Such disks might be expected to form stars rather than accrete onto the black hole. The arguments leading to this conclusion are reviewed. Conversion of a part of the gas into high-mass stars or stellar-mass black holes, and the release of energy in these objects by fusion or accretion, may help to stabilize the remaining gas. If the disk extends beyond a parsec, however, more energy is probably required for stability than is available by turning half the gas into high-mass stars. Small black holes are perhaps marginally viable energy sources, with important implications (not pursued here) for the QSO spectral energy distribution, the metallicity of the gas, microlensing of QSO disks, and perhaps gravitational-wave searches. Other possible palliatives for selfgravity include accretion driven by nonviscous torques that allow near-sonic accretion speeds and hence lower surface densities for a given mass accretion rate. All such modes of accretion face major theoretical difficulties, and in any case merely postpone selfgravity. Alternatively, thin disks may not exist beyond a thousand Schwarzshild radii or so (0.01 parsec), in which case QSOs must be fueled by gas with small specific angular momentum.

Key words: accretion disks—gravitation—quasars: general

1 INTRODUCTION

The outer parts of steady, geometrically thin, optically thick, viscously-driven accretion disks around QSOs are predicted to be selfgravitating: that is, the Toomre stability parameter Q [eq. (5)] is much less than unity; equivalently, the midplane density is much greater than the Roche value (Shlosman & Begelman 1987). As shown in §2 for a standard α disk, Q falls to unity at a distance $r \sim 10^{-2}$ pc from a central black hole of mass $10^8 M_\odot$ accreting at the Eddington rate. The mass of the disk within this radius is much less than that of the black hole and hence able to fuel QSO activity for only a fraction of the black-hole growth time, $t_{\rm Edd} \approx 5 \times 10^7 \varepsilon_{0.1}^{-1}$ yr. This is one reason to suspect that QSO disks are larger than 10^{-2} pc, but there is also indirect observational evidence. The spectral energy distribution of typical QSOs shows a strong peak in the rest-frame infrared, which is often interpreted as reprocessing of light from the central source by an extended warped disk (Sanders et al. 1989). Nearby Seyferts and other active galaxies, thought to be low-luminosity analogs of QSOs, sometimes have resolvable nuclear disks. Most spectacularly, VLBI observations of maser emission in NGC 4258 and NGC 1068 indicate disks on parsec scales and have been used to measure the mass of their central black holes (Nakai et al. 1993; Greenhill et al. 1995; Greenhill & Gwinn 1997).

At least on larger galactic scales, it is believed that strong self-gravity leads to star formation (Martin & Kennicutt 2001). If that is true of QSO disks, there is a danger that most of the gas would form stars, leaving little to fuel the QSO. Several theoretical attempts have been made to modify the standard α-disk model so as to extend gaseous QSO disks self-consistently into the selfgravitating regime. The options include driving accretion with global bars or density waves so as to increase the radial velocity and lower the surface density needed to sustain a given mass accretion rate (Shlosman & Begelman 1989); heating or mechanically stirring the disk with embedded stars or black holes so as to raise the temperature and lower the density of the gas (Collin & Zahn 1999a,b); or allowing the disk to fragment into colliding gas clumps whose epicyclic motions stabilize it against selfgravity, at least on scales larger than the clumps themselves (Kumar 1999).

Our purpose is to re-examine selfgravity in QSO accretion disks with an emphasis on dynamical and energetic constraints. These constraints are most severe for massive and luminous systems, so our interest is in black-hole masses $M \gtrsim 10^8 M_\odot$ and luminosities close to the Eddington limit.

The plan of the paper is as follows. §2 reviews selfgravity in steady α disks, including irradiation from the central source (§2.2), and enhancements to α by local gravitationally-driven turbulence in a moderately selfgravitating disk (§2.3). §3 considers disks that are stabilized by additional heating beyond that due to the dissipation of orbital energy, but in which angular momentum continues to be transported by an α viscosity. The effective temperature of such disks falls more slowly than the standard relation $T_{\rm eff} \propto r^{-3/4}$; in contrast to the usual situation, the total energy released per unit mass accreted is strongly dependent on the outer radius of the disk (3). §4 briefly explores various alternatives to viscous thin-disk accretion. Schemes that involve thin disk but invoke faster-than-viscous angular-momentum transport include global spiral waves (§4.5) and magnetized disk winds (§4.3). In every such case, we argue that the accretion velocity is bounded by the sound speed. Other alternatives explored in §4 are less like thin disks: quasi-spherical flows (§4.1), collisional star clusters (§4.2) and clumpy disks (§4.6).

All of these alternatives to the standard alpha disk face severe theoretical difficulties, or else seem unlikely to permit a centrifugally supported accretion flow beyond $\sim 0.1\,$pc. We conclude in §5 that the gas probably does not circularize beyond this radius and must be supplied to the nucleus with low specific angular momentum.

2 STEADY ALPHA DISKS

For the time being, we assume steady accretion and neglect winds, so that the mass accretion rate is constant with radius, and advection of angular momentum is balanced by viscous torque.

1. The mass accretion rate and viscosity parameter are related by

$$\dot{M} = 3\pi\alpha\beta^b c_{\rm s}^2 \Omega^{-1} \Sigma, \tag{1}$$

in which $c_{\rm s} = \sqrt{p/\rho}$ is the isothermal sound speed at the disk midplane, $\Omega = (GM/r^3)^{1/2}$ is the orbital angular velocity, Σ is the surface mass density, and $\beta \equiv p_{\rm gas}/p$ is the ratio of gas pressure to total pressure at the midplane. The dimensionless Shakura & Sunyaev (1973) viscosity parameter α is not necessarily constant but presumably $\lesssim 1$. The mechanism likely responsible for the viscosity of most disks is magnetorotationally-driven turbulence, for which simulations indicate $\alpha = 10^{-3} - 10^{-1}$ (Balbus & Hawley 1998). As discussed below, where the disk is selfgravitating, α may be as large as ~ 0.3. Therefore, we often write $\alpha_{0.01} \equiv 10^2\alpha$ or $\alpha_{0.3} \equiv \alpha/0.3$. The parameter b is a switch that determines whether the viscosity is proportional to gas pressure ($b = 1$) or total pressure ($b = 0$). It is well known that radiation-pressure-dominated regions of α disks have viscous instabilities in the latter case (Lightman & Eardley 1974), but it is possible that $b = 0$ in an average sense. Although the average surface density $\bar{\Sigma}$ would be lower and the corresponding $\bar{Q}$ higher for $b = 0$ than for $b = 1$, the viscous instability is expected to produce overdense rings in which these trends may well be reversed.

2. If the disk were heated by viscous dissipation only, then its surface temperature would be

$$T_{\rm eff} \quad = \quad \left(\frac{3}{8\pi\sigma} \frac{GM\dot{M}}{r^3} \right)^{1/4} \tag{2}$$

$$\approx \quad 2.9\times 10^3\,(M_8\dot{M}_\odot)^{1/4}(r/10^{-2}\,\mathrm{pc})^{-3/4}\ \ \mathrm{K},$$
$$\approx \quad 6.2\times 10^5\left(\frac{l_\mathrm{E}}{\varepsilon_{0.1}M_8}\right)^{1/4}\left(\frac{r}{R_\mathrm{S}}\right)^{-3/4}\ \ \mathrm{K}.$$

We have introduced the abbreviations $M_8 \equiv M/(10^8 M_\odot)$, $\dot{M}_\odot \equiv \dot{M}/(1\,M_\odot\ \mathrm{yr}^{-1})$, dimensionless luminosity $l_\mathrm{E} \equiv L/L_\mathrm{E}$, radiative efficiency $\varepsilon \equiv L/\dot{M}c^2 \equiv 0.1\varepsilon_{0.1}$, and Schwarzshild radius $R_\mathrm{S} = 2GM/c^2 \approx 10^{-5}M_8$ pc. The mass accretion rate can then be written as

$$\dot{M} = \frac{4\pi GM}{\kappa_{\mathrm{e.s.}}c}\,\frac{l_\mathrm{E}}{\varepsilon} \approx 2.2\,\varepsilon_{0.1}^{-1}\,l_\mathrm{E}\,M_8\,\dot{M}_\odot, \tag{3}$$

where $\kappa_{\mathrm{e.s.}} \approx 0.4\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ is the electron-scattering opacity.

3. If viscous dissipation occurs mostly near the midplane, and vertical transport of heat is by radiative diffusion, then the midplane and surface temperatures are related approximately by

$$T^4 \approx \frac{\kappa\Sigma}{2}T_\mathrm{eff}^4. \tag{4}$$

Simulations of magnetorotational turbulence indicate that the vertical scale height of the turbulent dissipation is larger than that of the gas density (Miller & Stone 2000). Hence eq. (4) may somewhat overestimate the midplane temperature, in which case selfgravity may extend to even smaller radii than estimated below.

2.1 Onset of selfgravity

The Toomre stability parameter for keplerian rotation is

$$Q = \frac{c_\mathrm{s}\Omega}{\pi G\Sigma} \approx \frac{\Omega^2}{2\pi G\rho}\,, \tag{5}$$

and local gravitational instability occurs where $Q < 1$. We have taken $\Sigma = 2h\rho$ with $h = c_s/\Omega$, which is approximately the disk half-thickness when $Q \gtrsim 1$. Combining eq. (5) with eq. (1), one obtains the simple relation

$$G\dot{M}Q = 3\alpha\beta^b c_\mathrm{s}^3\,. \tag{6}$$

For future reference, note that in a flat rotation curve, $v_\mathrm{circ} = r\Omega =$ constant, the numerical factor 3 is replaced by $2\sqrt{2}$: a distinction too fine to matter.

To see why selfgravity is inevitable, consider the radiation-pressure-dominated case, $\beta \ll 1$, so that the isothermal sound speed $c_s^2 = 4\sigma T^4/3c\rho$. From eqs. (2) & (4), $T^4 = 3\kappa\Sigma\Omega^2\dot{M}/16\pi\sigma$. Eliminating T between these relations and using $\Sigma = 2c_s\rho/\Omega$ leads to

$$c_s = \frac{\kappa\Omega\dot{M}}{2\pi c} = \frac{l_\mathrm{E}\kappa}{\varepsilon\kappa_{\mathrm{e.s.}}}\,\Omega R_\mathrm{S} \qquad (\beta \ll 1), \tag{7}$$

whence $h = (l_\mathrm{E}/\varepsilon)R_\mathrm{S}$ =constant for $\kappa = \kappa_{\mathrm{e.s.}}$. Using this for c_s and eq. (3) for $\dot{M}$ in eq. (6) leads to

$$Q = \frac{3\alpha\beta^b}{8\pi\sqrt{2}}\left(\frac{l_\mathrm{E}}{\varepsilon}\right)^2\hat{\kappa}^3\left(\frac{\kappa_{\mathrm{e.s.}}c^4}{G^2M}\right)\left(\frac{R_\mathrm{S}}{r}\right)^{9/2}, \qquad (\beta \ll 1), \tag{8}$$

in which $\hat{\kappa} \equiv \kappa/\kappa_{\mathrm{e.s.}}$. Hence disks around more massive black holes are more prone to selfgravity. For $b = 0$, there is a characteristic mass above which an Eddington-limited disk would be selfgravitating even at its inner edge. This mass is enormous ($\sim 10^{19}M_\odot$ for $\alpha_{0.01} = \varepsilon_{0.1} = 1$), but even for realistic black-hole masses, $Q < 1$ at radii greater than

$$r_{\mathrm{s.g.}} \approx 2.1\times 10^3\left(\frac{\alpha_{0.3}\beta^b l_\mathrm{E}^2\hat{\kappa}^3}{\varepsilon_{0.1}^2 M_8}\right)^{2/9} R_\mathrm{S} \qquad (\beta \ll 1). \tag{9}$$

Notice that α has been scaled to 0.3 rather than 10^{-2} because of the expected enhancement by gravitational turbulence (see below).

We now examine the assumption that $\beta \ll 1$. It follows from eqs. (3), (5), and (6) that

$$\begin{aligned}\frac{\beta^{1-(b/2)}}{(1-\beta)^{1/4}} &= 2^{-5/2}\left(\frac{3}{\pi}\right)^{3/4}\left(\frac{\varepsilon\alpha}{l_{\rm E}}\right)^{1/2} Q^{-3/4}\left(\frac{k_{\rm B}^4 c^9 \kappa_{\rm e.s.}{}^2}{m^4 G^7 \sigma}\right)^{1/4}\left(\frac{r}{R_{\rm S}}\right)^{-3/4} M^{-1} \\ &\approx 0.39\left(\frac{\varepsilon_{0.1}\alpha_{0.3}}{l_{\rm E}}\right)^{1/2} Q^{-3/4}\left(\frac{r}{2100R_{\rm S}}\right)^{-3/4} M_8^{-1}.\end{aligned} \tag{10}$$

Eq. (10) does not depend upon eqs. (2) & (4): that is, it does not assume that radiative losses are balance by viscous dissipation. (The opacity $\kappa_{\rm e.s.}$ enters the formula only because it is used in the definition of the Eddington ratio $l_{\rm E}$). Hence eq. (10) remains true even when auxiliary sources of heating are present in the disk. For viscously heated, radiatively cooled disks, β is given by eq. (A3) in the Appendix.

2.2 Irradiation

Although flared or warped outer regions can be warmed somewhat by irradiation from the inner parts (where most of the total disk luminosity originates), we now show this effect is not enough to stabilize the disk at $r \sim r_{\rm s.g.}$.

To do so, irradiation must raise the midplane temperature T, not just the surface temperature $T_{\rm eff}$, substantially. As a result, the disk will be nearly isothermal from surface to midplane. The vertical pressure *gradient*, and therefore the disk thickness, will be dominated by the gas even if $\beta \ll 1$: that is, $h^2 \approx p_{\rm gas}/\rho\Omega^2$. If the effective viscosity derives from local magnetorotational turbulence, then it probably scales more directly with thickness than with sound speed since MRI instabilities do not require compression. Then in all relevant respects the disk behaves as if $\beta = 1$ and $c_s^2 \to p_{\rm gas}/\rho = k_{\rm B}T/m$. From eq. (6), it then follows that the minimum temperature that the irradiation must provide to ensure gravitational stability is

$$T_{Q=\beta=1} = \frac{m}{k_{\rm B}}\left(\frac{G\dot{M}}{3\alpha}\right)^{2/3} \approx 5.6\times10^4\left(\frac{l_{\rm E}M_8}{\alpha_{0.3}\varepsilon_{0.1}}\right)^{2/3}\ {\rm K}. \tag{11}$$

(To be conservative, we have scaled to $\alpha = 0.3$; for $\alpha = 0.01$, this temperature would be an order of magnitude larger). On the other hand, the temperature of the disk in equilibrium with radiation from the central source is

$$T_{\rm eq} = \left(\frac{c^5 l_{\rm E}\cos\theta}{4\sigma\kappa_{\rm e.s.}GM}\right)^{1/4}\left(\frac{r}{R_{\rm S}}\right)^{-1/2} \approx 3.8\times10^5\left(\frac{l_{\rm E}\cos\theta}{M_8}\right)^{1/4}\left(\frac{r}{R_{\rm S}}\right)^{-1/2}\ {\rm K}, \tag{12}$$

where θ is the angle between the local normal to the disk and the radial rays. For a flat disk, $\cos\theta \sim \max(h, R_{\rm S})/r$ at $r \gg R_{\rm S}$, but even for a severely warped disk with $\cos\theta \sim 1$, $T_{\rm eq} \ll T_{Q=\beta=1}$ at $r/R_{\rm S} \gg 44(\alpha_{0.3}\varepsilon_{0.1})^{4/3} l_{\rm E}^{-5/6} M_8^{-11/6}$. Since this last inequality certainly holds at all $r \geqslant r_{\rm s.g.}$, we conclude that irradiation is not important for the selfgravity of the disk.

2.3 Local gravitational turbulence

It has occasionally been proposed that a partially selfgravitating disk may transport angular momentum by spiral waves (*e.g.* Cameron 1978; Paczyński 1978). Others have suggested that gravitational instabilities are intrinsically global and therefore not reducible to a local viscosity (Balbus & Papaloizou 1999). It may be that both opinions are sometimes correct, depending upon the ratio of disk thickness to radius, since the most unstable wavelength in a $Q = 1$ disk is $\sim h$. We consider local gravitational turbulence here, since it can be accommodated within the α model; a (much larger) upper bound on angular momentum transport by global spirals is given in §4.5.

By careful two-dimensional simulations, Gammie (2001) finds that gravitationally-driven turbulence can be local and can support α approaching unity; in fact, in the absence of any other viscosity mechanism, his disks self-regulate themselves so that

$$\alpha_{\rm grav} \approx \frac{1}{(\gamma_{\rm 2D}-1)\Omega t_{\rm th}}, \tag{13}$$

where $t_{\rm th} \equiv \Sigma k_{\rm B} T/\sigma T_{\rm eff}^4$ is the local thermal timescale, and $\gamma_{\rm 2D}$ is the 2D adiabatic index:

$$\gamma_{\rm 2D} = \left.\frac{\partial}{\partial \log \Sigma}\right)_S \log\left[\int\limits_{-\infty}^{\infty} p(r,\theta,z)\,dz\right].$$

[Our eq. (13) differs from Gammie by a factor $\gamma_{\rm 2D}$ because we define α in terms of isothermal rather than adiabatic sound speed.] Gammie finds that self-regulation fails and the disk fragments if $\Omega t_{\rm th} \lesssim 0.3$ for $\gamma_{\rm 2D} = 2$. The latter result is not entirely surprising: in the absence of pressure support, the natural timescale for collapse is $t_{\rm dyn} = \Omega^{-1}$, and one would not expect collapse to be prevented by thermal energy unless $t_{\rm th} > t_{\rm dyn}$. The converse statement, that fragmentation can be indefinitely postponed if $\Omega t_{\rm th} > 0.3$, is not at all obvious but is supported by Gammie's simulations.

We are not aware of any direct numerical simulations of disks that are unstable to both gravitational and magnetorotational modes at the same radii, as is likely to be the case for QSO disks (Menou & Quataert 2001). It would be interesting to explore this, as there might be a synergy between the two instabilities.

QSO disks are expected to be very thin in the regions of interest to us, so that Gammie's results may be applicable. Eq. (7) implies $h = l_{\rm E} R_{\rm S}/6\varepsilon$ in a radiation-pressure dominated disk, and hence $h/r \lesssim 0.003$ at the innermost radius where $Q \sim 1$ [eq. (9)]. We expect that from this radius outwards, α will rise smoothly from the value supported by magnetorotational turbulence (perhaps $\alpha_{\rm m.h.d.} \sim 10^{-2}$) up to the maximum allowed by gravitational turbulence, $\alpha_{\rm grav,max} \sim 0.3$, in such a way that $Q \approx$constant. It follows from eq. (8) that the ratio of outer to inner radii of this region is rather modest: $\sim (\alpha_{\rm grav,max}/\alpha_{\rm m.h.d.})^{2/9} \sim 2$. At still larger radii, additional sources of energy are required in order to prevent catastrophic fragmentation.

3 CONSTANT-Q DISKS

In this section, we postulate that some heating process arises that maintains the disk on threshold of gravitational instability, *i.e.* at a constant $Q \approx 1$, and work out some consequences for the structure of the disk. In §3.1, we discuss the radial dependence of midplane temperature, density, and β [already given by eq. (10)]. In §3.2, we estimate the minimal amount of energy in excess of viscous dissipation that must be supplied to maintain constant Q.

For our immediate purposes, the details of the heating process are not important as long as it provides a stable feedback that regulates Q. Angular momentum is assumed still to be viscous, though perhaps with an enhanced α in eq. (1). The effective temperature of the disk may no longer be determined by equation (2), however, because viscous dissipation may not dominate the thermal energy budget.

Although they may seem artificial, these assumptions are probably appropriate for the disks of spiral galaxies, and for the local interstellar medium in particular. The local Galactic magnetic field is consistent with simulations of magnetorotational turbulence: *viz.* a somewhat sub-equipartition strength, a predominantly toroidal orientation, and fluctuations comparable to the mean (Brandenburg et al. 1995). It is plausible therefore that there is a nonzero average magnetic stress $\langle B_r B_\theta\rangle/4\pi = \alpha p_{\rm gas}$ that systematically transfers angular momentum outwards (Sellwood & Balbus 1999). Taking $p_{\rm gas}/k_{\rm B} \approx 2000$ cm^{-3} K^{-1}, $\rho \approx 0.3\, m_H$ cm^{-3} (Spitzer 1978), and circular velocity $V_0 \approx 200$ km s^{-1} (Binney & Merrifield 1998), the implied radial drift velocity $v_r \approx -\alpha p_{\rm gas}/\rho V_0 \approx -0.3\,\alpha$ km s^{-1} is small enough to have escaped detection. Perhaps coincidentally, eq. (6) predicts $\dot M \approx 3\times10^{-2}\alpha_{0.1} Q^{-1} \dot M_\odot\,{\rm yr}^{-1}$, about half the Eddington rate for the Galaxy's $2.5\times10^6 M_\odot$ central black hole. The implied viscous heating rate $\alpha p_{\rm gas}\Omega_0 \approx 2\times10^{-29}\,\alpha_{0.1}\,{\rm erg\,cm^{-3}}$ is negligible compared to the inferred radiative cooling rate of the gas, $\sim 2\times10^{-26}{\rm erg\,cm^{-3}}$ (Spitzer 1978). Presumably, the temperature of the ISM is maintained by stars. An important difference between the the local ISM and QSO disks is that the former is very optically thin, especially to absorption, which means that the energy input from stars is inefficiently radiated.

3.1 Density and temperature

The midplane density in a constant-Q disk follows from eq. (5):

$$\rho = \frac{M}{2\pi Q r^3} = 1.2 M_8^{-2} Q^{-1} \left(\frac{R_{\rm S}}{r}\right)^3 \ {\rm g\,cm^{-3}}\,, \tag{14}$$

so that the density at $r_{\rm s.g.}$ [eq. (9)] is $\sim 10^{-8} M_8^{-4/3}\,{\rm g\,cm^{-3}}$. The ratio $\beta/(1-\beta) = 4\sigma T^3/3c\rho$ is determined by eq. (10). The temperature itself is (for $\beta \ll 1$)

$$\begin{aligned} T &= 2^{-7/6}\left(\frac{3}{\pi}\right)^{1/12}\left(\frac{l_{\rm E}}{\alpha\varepsilon Q^{1/2}}\right)^{1/6} c^{19/12} G^{-5/12} \sigma^{-1/4} \kappa_{\rm e.s.}{}^{-1/6}\, M^{-1/3}\left(\frac{R_{\rm S}}{r}\right)^{3/4} \\ &\approx 6.9\times 10^6 \left(\frac{l_{\rm E}}{\alpha_{0.3}\varepsilon_{0.1} Q^{1/2}}\right)^{1/6} M_8^{-1/3}\left(\frac{R_{\rm S}}{r}\right)^{3/4} \ {\rm K} \qquad \text{if } b=0\,; \end{aligned} \tag{15}$$

$$\begin{aligned} T &= 2^{-1/3}\left(\frac{\pi}{3}\right)^{1/6}\left(\frac{l_{\rm E} Q^{1/2}}{\alpha\varepsilon}\right)^{1/3} c^{5/6} G^{1/6} \sigma^{-1/6} \kappa_{\rm e.s.}{}^{-1/3} (k_{\rm B}/m)^{-1/3}\left(\frac{R_{\rm S}}{r}\right)^{1/2} \\ &\approx 1.4\times 10^6 \left(\frac{l_{\rm E} Q^{1/2}}{\alpha_{0.3}\varepsilon_{0.1}}\right)^{1/3}\left(\frac{R_{\rm S}}{r}\right)^{1/2} \ {\rm K} \qquad \text{if } b=1\,. \end{aligned} \tag{16}$$

The surface density is

$$\begin{aligned} \Sigma &= \left(\frac{l_{\rm E}\, c^8}{12\pi^2\sqrt{2}\,\varepsilon\alpha Q^2\, G^4 \kappa_{\rm e.s.} M^2}\right)^{1/3}\left(\frac{R_{\rm S}}{r}\right)^{3/2} \\ &\approx 7.4\times 10^9 \left(\frac{l_{\rm E}}{\alpha_{0.3}\varepsilon_{0.1} Q^2}\right)^{1/3} M_8^{-2/3}\left(\frac{R_{\rm S}}{r}\right)^{3/2} \ {\rm g\ cm^{-2}} \qquad \text{if } b=0; \end{aligned} \tag{17}$$

$$\begin{aligned} \Sigma &= \frac{2}{3}\left(\frac{2 l_{\rm E}\, m}{\varepsilon\alpha\, k_{\rm B}\kappa_{\rm e.s.}}\right)^{2/3}\left(\frac{3\,\sigma c^7}{\pi Q\, G}\right)^{1/6}\left(\frac{R_{\rm S}}{r}\right) \\ &\approx 5.4\times 10^8\, Q^{-1/6}\left(\frac{l_{\rm E}}{\alpha_{0.3}\varepsilon_{0.1}}\right)^{2/3}\left(\frac{R_{\rm S}}{r}\right) {\rm g\ cm^{-2}} \qquad \text{if } b=1. \end{aligned} \tag{18}$$

For the nominal values of the parameters shown, Σ would fall to $1\,{\rm g\ cm^{-2}}$ at $40\,{\rm pc}$ for $b=0$ and $5\,{\rm kpc}$ for $b=1$; however, in the latter case h/r rises more rapidly with r and is already ~ 0.5 at $r = 1\,{\rm pc}$.

Beyond $10^4 - 10^5 R_{\rm S} \approx 0.1 - 1\, M_8\,{\rm pc}$, the above formulae predict $T \lesssim 5000$ K, so that the opacity $\kappa \ll \kappa_{\rm e.s.}$ and the disk becomes optically thin. (This assumes $M_8 = l_{\rm E} = \varepsilon_{0.1} = \alpha_{0.3} = 1$. Dust will raise opacity again at $T \lesssim 1700\,{\rm K}$.) The disk must then be supported by gas pressure, notwithstanding eq. (10) which presumes that the radiation is trapped. But at $\beta = 1$, the minimum temperature for gravitational stability is (11), which is about an order of magnitude larger than predicted by the formulae above. So in a marginally gravitationally stable disk, there must be an extended region where the temperature adjusts itself within a limited range ($5000\,{\rm K} \lesssim T \lesssim 10^4\,{\rm K}$) so that the disk is marginally optically thin. At the low densities relevant here, the maximum opacity is $\kappa_{\rm max} \approx 10\kappa_{\rm e.s.}$ and is achieved at $T \approx 10^4$ K (Kurucz 1992; Keady & Kilcrease 2000). Hence the outer edge of the region in question should end at $\Sigma \approx \kappa_{\rm max}^{-1} \approx 0.3\,{\rm cm^2\,g^{-1}}$ which occurs (assuming $Q = 1$ and $T = 10^4$ K) at

$$r_{\rm thin} \approx \frac{c_{\rm s}\kappa_{\rm max} v_{\rm circ}}{\pi G} \approx 170 M_8^{0.24}\ {\rm pc}, \tag{19}$$

provided of course that the disk extends to such large radii. We have taken the circular velocity as $\sigma_{\rm bulge}\sqrt{2} \propto M^{0.24}$ [see eqs. (30)&(31)] rather than $\sqrt{GM/r}$, since $r_{\rm thin}$ lies outside the black hole's sphere of influence [eq. (23)].

The parts of the disk beyond $r_{\rm thin}$ would radiate predominantly in optical emission lines with velocity widths $\sim \sigma_{\rm bulge}$, so that they might be identified with the QSO narrow-line region. However, we will now see that the energy required to maintain the disk at constant $\dot{M}$ and $Q \gtrsim 1$ all the way out to $r_{\rm thin}$ would be prohibitive.

3.2 Energetics

We define a local disk efficiency by

$$\varepsilon'(r) \equiv \frac{4\pi r^2\sigma T_{\rm eff}^4}{3\dot{M}c^2}. \tag{20}$$

For a viscously heated disk in newtonian gravity, $\varepsilon' c^2$ reduces to the local binding energy per unit mass, $GM/2r = c^2 R_{\rm S}/4r$ [see eq. (2)], which is largest at small radii; given a torque-free inner boundary at $r = r_{\rm min}$, the global efficiency is $\varepsilon = \varepsilon'(r_{\rm min})$. But the constant-$Q$ disks require additional energy inputs, so that ε' is generally larger than $R_{\rm S}/4r$ and tends in fact to *increase* with radius.

Using the results of §3.1 for the midplane temperature T, and assuming that vertical radiative transport obeys eq. (4), we find that

$$\begin{aligned}\varepsilon'(r) &= \frac{\pi^{1/3}}{2^{7/6}3^{1/3}}\,\hat\kappa^{-1}\left(\frac{Q\varepsilon^2}{\alpha l^2}\,\frac{GR_{\rm S}}{c^2\kappa_{\rm e.s.}}\right)^{1/3}\left(\frac{r}{R_{\rm S}}\right)^{1/2}\\ &\approx 8.2\times10^{-4}\left(\frac{Q\varepsilon_{0.1}^2}{\alpha_{0.3}l_{\rm E}^2}\right)^{1/3}\hat\kappa^{-1}M_8^{1/3}\left(\frac{r}{10^5R_{\rm S}}\right)^{1/2}\qquad \text{if } b=0;\end{aligned} \tag{21}$$

$$\begin{aligned}&= \frac{\pi^{5/6}}{2^{1/6}3^{5/6}}\,\frac{\kappa_{\rm e.s.}^{1/3}}{\kappa}\,\frac{Q^{5/6}\varepsilon^{1/3}}{\alpha^{2/3}l_{\rm E}^{1/3}}\left(\frac{G}{c}\right)^{5/6}\left(\frac{m}{k_{\rm B}}\right)^{2/3}\sigma^{1/6}\,r\\ &\approx 1.1\times10^{-2}\left(\frac{Q^{5/2}\varepsilon_{0.1}}{\alpha_{0.3}^2 l_{\rm E}}\right)^{1/3}\hat\kappa^{-1}M_8\left(\frac{r}{10^5R_{\rm S}}\right)\qquad \text{if } b=1.\end{aligned} \tag{22}$$

It would appear from these relations that $\varepsilon' \to \infty$ as $\kappa \to 0$, but they are not valid when the disk is optically thin. If the absorption optical depth $\tau = \kappa\Sigma < 1$, then $T_{\rm eff}^4 \approx \tau T^4$ rather than T^4/τ. So in fact $\varepsilon' \to 0$ as $\kappa \to 0$. Hence the largest radius at which eqs. (21) or (22) is valid is the smallest among the radii $r_{\rm rthin}, R_{\rm N}$, and $r_{\rm out}$, where the first is the radius at which a $Q = 1$ disk would become optically thin [eq. (19)], the second is the radius within which the black hole dominates the potential,

$$R_{\rm N} \equiv \frac{GM}{2\sigma_{\rm bulge}^2} \approx (6\pm1)\,M_8^{0.5\pm0.1}\ {\rm pc}\,, \tag{23}$$

and the last is the actual outer edge of disk, which depends upon the angular momentum of the gas supplied to it. The final expression in eq. (23) has used the recently reported $M - \sigma_{\rm bulge}$ relation (Gebhardt et al. 2000; Merritt & Ferrarese 2001, see §4.2 below). Actually, eq. (23) probably underestimates the distance to which the black hole dominates the circular velocity, because the density profile in the cusp of bright bulges is considerably shallower than r^{-2}.

At $r = 10$ pc $\approx 10^6\,M_8^{-1}R_{\rm S}$, equations (21) and (22) imply $\varepsilon' \approx 0.003\,M_8^{-1/6}$ and $\varepsilon' \approx 0.1$, respectively. The former efficiency is barely compatible with thermonuclear burning even if all of the disk gas is processed through high-mass stars. The latter is unsustainable by stars and in fact comparable to the efficiency of the central engine.

4 ALTERNATIVES TO THIN DISKS

As shown in §2, a geometrically thin, optically thick accretion disk in a typical high-luminosity QSO would be selfgravitating at radii $r \gtrsim 10^2 - 10^3 R_{\rm S}$, or $10^{-3} - 10^{-2}$ pc, where $R_{\rm S} = 2GM/c^2$ is the Schwarzschild radius of the central black hole. Large global radiative efficiency, $\varepsilon \gtrsim 0.1$, probably requires a thin disk near $R_{\rm S}$, but it does not much depend upon the nature of the flow at large radius. So in this section, we consider whether the selfgravitating part of the disk can be replaced by some other form of accretion.

4.1 Hot, quasi-spherical flows

These have been proposed as models for accretion at very low accretion rate ($\dot{M}$) and low radiative efficiency $\varepsilon \equiv L/\dot{M}c^2$ (Rees et al. 1982; Ichimaru 1977; Narayan & Yi 1994). At sufficiently low density, the gas cooling time and ion-electron thermal equilibration time are longer than the accretion time, so that the ion temperature (T_i) is approximately virial and the thickness of the disk is comparable to its radius. One supposes that angular-momentum transport is efficient (viscosity parameter $\alpha \sim 1$), or that the angular momentum is very small to begin with, so that the accretion velocity (v_r) is comparable to the free-fall velocity. The large v_r and T_i combine with the low $\dot{M}$ to make the density low, as required. It is unclear just how low $\dot{M}$ must be for this mode of accretion to sustain itself, because angular-momentum transport is not well understood and collisionless processes may enhance the thermal coupling of ions and electrons.

A quasi-spherical flow is not viable when the luminosity is close to the Eddington limit, however, because of inverse-Compton cooling. For spherical free fall onto a source of luminosity $L = l_{\rm E}L_{\rm Edd}$, the inverse-Compton cooling rate of free electrons is

$$t_{\rm C}^{-1} \approx \frac{2\sigma_{\rm T}L}{3\pi r^2 m_e c^2} = \frac{8}{9}\frac{m_p}{m_e} l_{\rm E}\left(\frac{R_{\rm S}}{r}\right)^{1/2} t_{\rm ff}^{-1}\,. \tag{24}$$

We have introduced the free-fall time of a radial parabolic orbit from radius r,

$$t_{\rm ff} \equiv \frac{2}{3}\left(\frac{r^3}{2GM}\right)^{1/2} \approx 2.2\times 10^{10}\, M_8^{-1/2} r_{\rm pc}^{3/2}\ {\rm s}, \tag{25}$$

in which $r_{\rm pc}$ is the distance from the source in parsecs and $M_8 \equiv M/(10^8 M_\odot)$. Actually, radiation pressure increases the free-fall time by $(1-l_{\rm E})^{-1/2}$, but this factor is neglected for simplicity. From eq. (24), $t_{\rm C} < t_{\rm ff}$ at $r < 25 l_{\rm E}^2 M_8$ pc. Hence the electrons assume the color temperature of the central source (or even less, see below): $T_{\rm C} \lesssim 10^5 M_8^{-1/4}$. This is much less than the virial temperature, $T_{\rm vir} \approx 10^7 M_8 r_{\rm pc}^{-1}$, at all radii of interest to the present paper. [1] The electron density is

$$n_e = \left(\frac{l_{\rm E}}{\varepsilon}\right)\frac{1}{3c\sigma_{\rm T}\,t_{\rm ff}}, \tag{26}$$

so that the electron-ion equilibration time due to Coulomb collisions alone (Spitzer 1978) is much shorter than the flow time:

$$\frac{t_{\rm eq}}{t_{\rm ff}} \approx 2\times 10^{-5}\varepsilon l_{\rm E}\left(\frac{T_e}{10^5\,{\rm K}}\right)^{3/2}\,.$$

Of course $T_e \sim 10^5$ K is the peak of the cooling curve (Spitzer 1978). If n_e obeys eq. (26), then radiative cooling is actually faster than inverse-Compton cooling:

$$\frac{t_{\rm rad}}{t_{\rm C}} \approx 6.\varepsilon\left(\frac{R_{\rm S}}{r}\right)^{1/2} \qquad (T_e = 10^5\,{\rm K}).$$

In short, both the ions and the electrons of a quasi-spherical flow onto a near-Eddington QSO cool in much less than a free-fall time. A thin disk will form unless the specific angular momentum of the flow is negligible.

4.2 Collisional stellar cluster

Dense stellar clusters have occasionally been nominated as precursors to QSO black holes, either by relativistic collapse (Zel'Dovich & Podurets 1966; Shapiro & Teukolsky 1985; Ebisuzaki et al. 2001) or by collisions among non-degenerate stars (Spitzer & Saslaw 1966; Rees 1978). Our interest, however, is in stellar collisions as the main source of fuel for an already very massive black hole. This has been studied by McMillan et al.

[1] The timescale for Poynting-Robertson drag is $\approx (m_p/m_e)t_{\rm C}$, hence $\gg t_{\rm ff}$ at $r \gg R_{\rm S}$, *i.e.*, the radiation field does not remove angular momentum from the gas fast enough to prevent it from circularizing.

(1981) and Illarionov & Romanova (1988), among others, who show that in order to supply a $10^8 M_\odot$ black hole at its Eddington rate, the velocity dispersion of such a cluster must be $\gtrsim 10^3\,\mathrm{km\ s^{-1}}$. To demonstrate the robustness of this conclusion, we will make some oversimplified but conservative estimates here.

In order to provide high radiative efficiency (ε), stars must be disrupted and their gaseous debris circularized before accretion. If $M_{\mathrm{b.h.}} \lesssim 10^8 M_\odot$, main-sequence stars scattered onto loss-cone orbits are likely to be tidally disrupted rather than swallowed whole (Hills 1975). About half of the tidal debris is unbound and promptly escapes from the black hole (Lacy et al. 1982; Evans & Kochanek 1989), and much of the remainder is likely to be swallowed at low radiative efficiency before the gas circularizes (Cannizzo et al. 1990; Ayal et al. 2000). We limit our discussion to $M_{\mathrm{b.h.}} \gtrsim 10^8 M_\odot$, as required for the most luminous QSOs, and assume that stars are disrupted by stellar collisions. We ignore loss-cone effects because stars swallowed whole do not contribute to the QSO luminosity.

The total collision rate among N stars forming a cluster of structural length a is

$$\dot{N} = C\,\bar{\sigma}^7\,N^2(GM)^{-3}R_*^2, \tag{27}$$

where C is a dimensionless coefficient, $\bar{\sigma}$ the root-mean-square velocity dispersion in one dimension averaged over the cluster, N the number of stars, and $R_* \sim R_\odot$ the radius of an individual star. M is the total mass that determines $\bar{\sigma}$ *via* the virial theorem, so that if m_* is the mass of individual stars and $M_* \equiv Nm_*$ then $M \approx M_{\mathrm{bh}} + \frac{1}{2}M_*$, the factor of 1/2 being needed to avoid double-counting the gravitational interactions among the stars. We take $4\pi R_*^2$ for the collision cross section; this allows for grazing collisions that probably would not disrupt the stars (Spitzer & Saslaw 1966), but it will lead to a conservative estimate of $\bar{\sigma}$. Gravitational focusing is unimportant at the high velocity dispersions relevant here.

A second relation for $\dot{N}$ follows by requiring the collisional debris to sustain the QSO at a fraction $l_{\mathrm{E}} \equiv L/L_{\mathrm{E}}$ of its Eddington luminosity:

$$2M_*\dot{N} = \frac{l_{\mathrm{E}}}{\varepsilon}\frac{4\pi GMm_p}{c\sigma_{\mathrm{T}}}. \tag{28}$$

Eliminating $\dot{N}$ between eqs. (27) and (28) leads to

$$\bar{\sigma}^7 = \frac{2\pi l_{\mathrm{E}}}{C\varepsilon}\,\frac{G^4 m_p}{\sigma_{\mathrm{T}}c}\,\frac{m_*}{R_*^2}\left(\frac{M_{\mathrm{b.h.}}M^3}{M_*^2}\right)\ .$$

The term in parentheses achieves its minimum value $27M_{\mathrm{b.h.}}^2/16$ at $M_* = 4M_{\mathrm{b.h.}}$.

The coefficient C depends upon the density profile $\rho(r)$ of the stellar cluster. It can be arbitrarily large if $\rho(r)$ is sufficiently steep, but then the collision rate is dominated by stars at small radius. These tightly-bound stars represent only a small fraction of the cluster mass and would be consumed in much less than the growth time of the black hole, unless their total mass ($M_{\mathrm{t.b.}}$) is $\gtrsim M_{\mathrm{b.h.}}$, in which case the tightly-bound population is substantially self-gravitating and we redefine $M_* \equiv M_{\mathrm{t.b.}}$. One might suppose that $M_{\mathrm{t.b.}} \ll M_{\mathrm{b.h.}}$ and that the tightly-bound stars were continuously replenished by two-body relaxation from a reservoir of more weakly bound stars; but the larger cluster would then have to expand to conserve energy, reducing the collision rate and the fueling of the QSO. For these reasons, we consider clusters for which C is dominated by stars near the half-mass radius. For example, if the stars have isotropically distributed orbits with a common semimajor axis a in a keplerian potential, then $\rho(r) \propto \sqrt{(2a/r)-1}$, $\bar{\sigma}^2 = GM/3a$, and $C \approx 196$. This leads to

$$\bar{\sigma} \approx 760\left(\frac{l_{\mathrm{E}}}{10\varepsilon}\right)^{1/7}\left(\frac{g_*}{g_\odot}\right)^{1/7} M_8^{2/7}\ \mathrm{km\ s^{-1}}, \tag{29}$$

where g_* is the stellar surface gravity. Because of the one-seventh root, the result is not very sensitive to our assumptions. A Plummer sphere of the same total mass, $M_* = 5\times 10^8 M_\odot$, yields $\bar{\sigma} \approx 730\,\mathrm{km\ s^{-1}}$.

Equation (29) can be compared with recently-discovered empirical relations between inactive black holes—presumably QSO relics—and their host bulges. Gebhardt et al. (2000) find

$$M_{\mathrm{b.h.}} = 1.2(\pm 0.2)\times 10^8\left(\frac{\sigma_{\mathrm{e}}}{200\,\mathrm{km\ s^{-1}}}\right)^{3.75\pm 0.3} M_\odot\,, \tag{30}$$

where σ_e is the line-of-sight velocity dispersion at one effective radius. Merritt & Ferrarese (2001) use the central velocity dispersion, which is usually little different from σ_e:

$$M_{\rm b.h.} = 1.30(\pm 0.36)\times 10^{8}\left(\frac{\sigma_c}{200\,{\rm km\ s^{-1}}}\right)^{4.72\pm 0.36} M_\odot\,. \tag{31}$$

The scaling exponent $d\log M/d\log\sigma = 3.5$ implied by eq. (29) is similar to the empirical ones (30) & (31), but the normalization is very different. Extrapolated to $700\,{\rm km\ s^{-1}}$, the empirical relations predict $M_{\rm b.h.} \gtrsim 10^{10} M_\odot$ instead of $10^8 M_\odot$.

Therefore, if QSOs were fueled by dense stellar clusters, these clusters must have been an order of magnitude more tightly bound than the surrounding bulge, and they would have been a dynamically distinct stellar component. There seems to be very little trace of this tightly-bound stellar population in present-day bulges. How would such a component form? A likely possibility is gaseous dissipation followed by star formation. As will be seen, accretion in a thin viscous disk may lead to just such a result. We have discussed fueling the QSO by a disk and fueling it by stellar collisions as though these were mutually exclusive possibilities, but perhaps the two occur in concert.

4.3 Wind-driven disks

In principle at least, a magnetized wind can remove angular momentum from a thin disk rather efficiently. Compared to a viscous disk of the same sound-speed (c_s) and accretion rate ($\dot M$), a wind-driven disk might have an accretion velocity that is larger by a factor $\sim r\Omega/c_s = r/h$. The surface density would be correspondingly reduced, as would the tendency toward self-gravity.

If viscous transport can be neglected, then under steady conditions,

$$\dot M_{\rm acc}\Omega r^2 = -\int\limits_{r_{\rm min}}^{r} \overline{B_z B_\phi} r'^2\, dr'. \tag{32}$$

The lefthand side is the angular-momentum flux carried inward by the accreting gas. The mass flux through the disk $\dot M_{\rm acc}$, is now a function of r even in a steady disk because of the mass loss through the wind, $\dot M_{\rm wind} \ll \dot M_{\rm acc}$. The righthand side of eq. (32) is the rate of loss of angular momentum by maxwell stresses exerted on both faces of the disk, assuming that B_ϕ is odd in z while B_z is even. The overbars denote an average over fluctuations in azimuth and time. The minus sign appears because we define $\dot M_{\rm acc} > 0$ for inflow. The corresponding equation for a viscous disk is

$$\dot M_{\rm acc}\Omega r^2 = -\frac{1}{2} r^2 \int\limits_{-h}^{h} \overline{B_r B_\phi}\, dz. \tag{33}$$

The magnetic part of the angular-momentum flux emerges through surface areas $4\pi(r^2 - r_{\rm min}^2)$ and $4\pi r h$ in eqs. (32) & (33), respectively. Since the former area is larger than the latter by a factor r/h, the field needed to drive a given $\dot M_{\rm acc}$ is smaller by h/r in a wind-driven disk than in a viscous disk.

If the effective viscosity of an accretion disk is magnetic, then the energy density of the field is at most in equipartition with the gas, *viz* $\beta_{\rm mag} \equiv 8\pi p_{\rm gas}/\mathbf{B}^2 \gtrsim 1$, because a super-equipartion field would shut off the crucial magnetorotational instability (Balbus & Hawley 1998). We now give two arguments to show that even in a wind-driven accretion disk, the field must also be at or below equipartion.

Vertical hydrostatic equilibrium requires

$$\left.\frac{B_r^2 + B_\phi^2 - B_z^2}{8\pi} + p_{\rm gas}\right|_{z=0}^{z=h} \approx \int\limits_0^h \rho\Omega^2 z\, dz\,.$$

We assume that the field has dipolar symmetry, so that B_z is approximately constant with z on the scale $h \ll r$, whereas the horizontal components vanish at the midplane. Since the righthand side above is positive, and since $p_{\rm gas}(z=h) \ll p_{\rm gas}(z=0)$, it follows that

$$\frac{B_r^2 + B_\phi^2}{8\pi}(z = \pm h) \lesssim p_{\rm gas}(z = 0). \tag{34}$$

Henceforth B_r and B_ϕ are evaluated at $z = h$, and $p_{\rm gas}$ at $z = 0$. As pointed out by Blandford & Payne (1982), in order that centrifugal force should drive the wind outward, the poloidal field lines must make an angle of at most 60° with the surface of the disk, so that $B_r/B_z \geqslant \sqrt{3}$. Therefore

$$\frac{\mathbf{B}^2}{8\pi} \leqslant \frac{4B_r^2 + B_\phi^2}{8\pi} \leqslant 4p_{\rm gas}.$$

The magnetic energy density is less important to us, however, than the azimuthal force per unit area on the disk:

$$\left|\frac{B_z B_\phi}{4\pi}\right| \leqslant \sqrt{3}\left|\frac{B_r B_\phi}{4\pi}\right| \leqslant \sqrt{3}\frac{B_r^2 + B_\phi^2}{8\pi} < \sqrt{3}\,p_{\rm gas}. \tag{35}$$

Using the inequality in eq. (32), together with $\dot{M}_{\rm acc} \equiv -2\pi\Sigma v_r$ and $p_{\rm gas} \approx \Sigma c_{\rm s}^2/2h$, we have

$$|v_r| \lesssim \frac{\sqrt{3}}{\Omega r^3 \Sigma}\int\limits_{r_{\rm min}}^{r} \frac{\Sigma c_{\rm s}^2}{h}(r')\, r'^{\,2}\, dr' \sim \sqrt{3}\frac{c_{\rm s}^2}{\Omega h}. \tag{36}$$

So the inflow could be marginally supersonic. For a viscous disk, on the other hand, $|v_r| \approx \alpha c_{\rm s}^2/\Omega r$, which is strongly subsonic as long as $\alpha \ll r/h$.

The angular momentum extracted from the face of the disk by the field must be carried off by the wind. It is problematic whether a strong wind can be launched from the disk (Ogilvie & Livio 2001), and whether rapid wind-driven accretion is stable (Cao & Spruit 2001). Apart from these difficulties, consideration of the magnetic flux,

$$\Phi(r) \equiv \int\limits_{r_{\rm min}}^{r} B_z(r', 0)\, 2\pi r' dr',$$

leads to important constraints on the field strength and accretion rate. Presumably $\Phi(r)$ should not change secularly. If B_z is predominantly of one sign, then the increase of $|\Phi|$ by advection must be balanced by diffusion of the lines through the inflowing gas. The drift velocity of the lines $v_{\rm drift} = -v_r \sim \eta_{\rm eff}/r$ if B_z varies on scales $\sim r$, where $\eta_{\rm eff}$ is the effective diffusivity of the gas. In significantly ionized disks, the microscopic diffusivity is negligible, so $\eta_{\rm eff}$ is due to turbulence, and one expects $\eta_{\rm eff} = \alpha_{\rm mag} c_{\rm s} h$ with $\alpha_{\rm mag} \lesssim 1$ for much the same reasons that the effective viscosity $\nu_{\rm eff} = \alpha_{\rm mag} c_{\rm s} h$ with $\alpha_{\rm visc} \lesssim 1$. Hence for a steady wind-driven disk threaded by net magnetic flux, $|v_r| \lesssim (\alpha_{\rm mag} h/r) c_{\rm s}$, which is probably very much less than the upper limit (36) and comparable to the accretion velocity of a viscous disk. Alternatively, the net flux could be essentially zero if B_z changes sign on scales $\ll r$. In the latter case, the higher flow speed (36) may be achievable. But so irregular a field would probably have to be sustained by dynamo action within the disk rather than inherited from whatever region supplies the accreting gas. This probably requires magnetorotational instability (henceforth MRI) and gives another argument for a sub-equipartion field.

To summarize this subsection, accretion driven by magnetized winds is even less well understood than viscous accretion but might allow substantially higher accretion velocities and lower surface densities, perhaps by factors up to $\sim r/\alpha h$.

4.4 Thin disks with strongly magnetized coronae

This is a variant of §4.3 in which most of the field lines are not open but re-attach to the disk at large distances $\Delta r \gg h$ (Galeev et al. 1979; Heyvaerts & Priest 1989). The vertical magnetic scale height is then $H \sim \Delta r \gg h$. The angular momentum flux carried through the corona,

$$\dot{J}_{\rm cor}(r) = -2\int\limits_{z=h}^{z=\infty} dr \int\limits_{0}^{2\pi} d\phi\, \frac{B_r B_\phi}{4\pi} \;\sim\; H\,\frac{B_r B_\phi}{4\pi}(r, h)\,,$$

can be larger than the flux within the gas layer by a factor $\sim H/h$, so that the effective value of α might be as large as r/h (for $H \sim r$) without fields exceeding equipartion.

The evidence for magnetized coronae that dominate angular-momentum transport is suggestive but inconclusive. Local simulations of MRI generally find that the scale height of the field exceeds that of the gas, but only by factors of order unity; they also find $\alpha \sim 10^{-2} - 10^{-1}$ rather than $\sim r/h$ (Brandenburg et al. 1995; Stone et al. 1996; Miller & Stone 2000). Possibly, H/h is limited by the fact that the smallest dimension of the computational domain is $\lesssim h$. Global simulations of MRI have been performed for relatively thick disks only, so that it is difficult to distinguish scalings with h from scalings with r (Matsumoto & Shibata 1997; Hawley 2000). Global simulations of *thin* disks in three dimensions may not be available for some time because of the very large numbers of grid cells needed to resolve both the disk and the corona. Merloni & Fabian (2001) argue that X-ray observations of accreting black holes (AGN and X-ray binaries) demand a strongly magnetized corona, at least in the innermost part of the disk. On the other hand, observations of eclipsing cataclysmic variables indicate that X-rays are emitted from the disk-star boundary layer rather than an extended corona (Mukai et al. 1997; Ramsay et al. 2001).

4.5 Global spiral waves

As is well known, a trailing m-armed spiral density wave

$$\Sigma(r,\theta) = \Sigma_0(r) \; + \Sigma_m(r)\cos\left(m\theta + \mu\ln r\right) \tag{37}$$

exerts an outward (positive) gravitational angular-momentum flux

$$\Gamma \approx \pi^2 G\Sigma_m^2 r^3\,\frac{m\mu}{|\mu|^3}$$

(Lynden-Bell & Kalnajs 1972). The above approximation is good for tightly-wrapped waves, $\mu \gg m \neq 0$, which carry relatively little flux for a given density contrast Σ_m/Σ_0. An exact formula for logarithmic spirals with $\Sigma_m \propto r^{-3/2}$ is

$$\Gamma = -\pi^2 G\Sigma_m^2 r^3\, m\frac{\partial}{\partial\mu}K(\mu,m), \tag{38}$$

where $K(\mu,m)$ is the Kalnajs function (Kalnajs 1971):

$$K(\mu,m) = \frac{1}{2}\left|\frac{\Gamma\left[\frac{1}{2}(m+i\mu\,+\frac{1}{2})\right]}{\Gamma\left[\frac{1}{2}(m+i\mu\,+\frac{3}{2})\right]}\right|^2 \qquad (\text{real }\mu). \tag{39}$$

With $\Sigma_m/\Sigma_0 = 1$, the largest ratio for which the surface density is everywhere positive, one finds that the torque is maximized at $m = 1$ and a pitch angle $\tan^{-1}(m/\mu) \approx 48^\circ\!.2$, so that

$$\Gamma_{\max} \approx 0.961\pi G\Sigma_0^2 r^3. \tag{40}$$

Suppose that the gravitational torque is balanced by the advection of angular momentum with the accreting gas. In other words, $\Gamma \approx \dot{M}\Omega r^2$, so that there is no secular change in the angular momentum within radius r. The gravitationally-driven accretion speed is then

$$|v_r| \leqslant \frac{\Gamma_{\max}}{2\pi r^3\Omega\Sigma_0} \approx 0.15 Q^{-1}\,c_{\rm s} \tag{41}$$

in which Q has been calculated from the azimuthal average of the surface density, Σ_0. In principle therefore, accretion may occur at a significant fraction of the sound speed. But the existence of a selfconsistent wave-driven flow has been assumed rather than proved. Nonlinear single-armed spirals in keplerian disks have been found by Lee & Goodman (1999), but only for weak self-gravity ($Q \gg 1$) and without dissipation or accretion.

In steady accretion onto a central mass that dominates the rotation curve, the advected angular-momentum flux $\Gamma = \dot{M}\Omega r^2 \propto r^{1/2}$, so $\Sigma \propto r^{-5/4}$ rather than $r^{-3/2}$. Presumably the slight change in power-law index does not change the results (40)-(41) much.

4.6 Clumpy disks

Can QSO disks persist at $Q \ll 1$ without fragmenting entirely into stars? This question is all the more urgent because of rather direct evidence for parsec-scale accretion disks in nearby AGN, if not QSOs, from VLBI observations of maser emission. If the nuclear disk of NGC 1068 is in a steady state, then the nuclear luminosity implies $Q \sim 10^{-3}$ at $r \sim 1\,\mathrm{pc}$ (Kumar 1999). NGC 4258 is much less luminous, and estimates of $\dot{M}$ range from $7 \times 10^{-5}\alpha_{0.1}\dot{M}_\odot$ based on modeling the maser emission itself (Neufeld & Maloney 1995), to $10^{-2}\dot{M}_\odot$ (Gammie et al. 1999; Kumar 1999) for an assumed central ADAF; at the former rate, $Q \sim 1$ at the outer edge of the masering region, $\sim 0.2\,\mathrm{pc}$ (Maoz 1995), while in the latter, $Q \sim 10^{-2}$.

Kumar (1999) has suggested a clumpy rather than smooth disk, in which accretion occurs by gravitational scattering and physical collisions among clumps rather than an α viscosity. These clumps are supposed to be gas clouds rather than fully formed stars, in order to provide appropriate conditions for maser amplification. Although the model deals with the stability and accretion rate of the clumpy disk as a whole, it does not ask what prevents the individual clumps from collapsing. In the application to NGC 1068, the masses, radii, and surface temperatures of the clumps are quoted as $M_\mathrm{c} \sim 10^3 M_\odot$, $R_\mathrm{c} \sim 0.1\,\mathrm{pc}$, and $T_{\mathrm{eff,c}} \approx 500\,\mathrm{K}$; the virial temperature implied by this mass and radius is $\sim 2000\,\mathrm{K}$. These clumps would be moderately optically thick at the wavelength corresponding to $T_{\mathrm{eff,c}}$. Their characteristic thermal time—the time required to radiate their binding energy—is

$$t_{\mathrm{th,c}} \approx \frac{GM_\mathrm{c}^2/R_\mathrm{c}}{4\pi R_\mathrm{c}^2 \sigma T_{\mathrm{eff,c}}^4} \sim 10^5\,\mathrm{s}\,,$$

hence many orders of magnitude less than the orbital time ($\Omega^{-1} \sim 10^{11}\,\mathrm{s}$ at $r \sim 1\,\mathrm{pc}$) or interclump collision time. Of course the surface temperatures of these objects are assumed to be maintained by irradiation from the central source, but without a fast-responding feedback mechanism, the thermal equilibrium is unstable: if a clump starts to contract, its surface temperature will rise, and collapse will proceed on the clump's internal free-fall timescale (since this is larger than $t_{\mathrm{th,c}}$).

To answer the question raised above, we therefore believe that no QSO accretion disk, whether smooth or clumpy, can persist at $Q \ll 1$.

5 DISCUSSION

Although the accretion velocity of a thin viscous disk is very strongly subsonic, with a Mach number $\mathcal{M} \equiv v_r/c_\mathrm{s} \approx \alpha h/r$, the discussion of §4 suggests that $\mathcal{M}$ might approach ~ 0.1 if accretion is driven by large-scale magnetic or gravitational fields rather than a local effective viscosity. Self-consistent solutions that achieve this bound would be very interesting to pursue.

Even at near-sonic accretion speeds, QSO disks become selfgravitating at radii less than a parsec if there is no important source of heating other than dissipation of orbital energy. Combining eqs. (2) & (4) with $\dot{M} = 2\pi r \Sigma \mathcal{M} c_\mathrm{s}$ instead of the viscous relation (1), and assuming that gas pressure dominates ($\beta \approx 1$), one finds that

$$\begin{aligned} Q &= 2^{-25/18}\pi^{-1}\mathcal{M}^{7/9}\hat{\kappa}^{2/9}\kappa_{\mathrm{e.s.}}^{7/9}G^{-1}\left(\frac{ck_\mathrm{B}}{\sigma^{1/4}m}\right)^{8/9} R_\mathrm{S}^{-11/9}\left(\frac{r}{R_\mathrm{S}}\right)^{-25/18} \\ &\approx 0.8\left(\frac{r}{10^4 R_\mathrm{S}}\right)^{-25/18}\left(\frac{\mathcal{M}_{0.1}^7\varepsilon_{0.1}^5\hat{\kappa}^2}{l_\mathrm{E}^5 M_8^{11}}\right)^{1/9} \\ &\approx 0.8\left(\frac{r}{0.1\,\mathrm{pc}}\right)^{-25/18}\left(\frac{\mathcal{M}_{0.1}^7\varepsilon_{0.1}^5\hat{\kappa}^2}{l_\mathrm{E}^5}\right)^{1/9} M_8^{1/6}\,. \end{aligned} \tag{42}$$

On the other hand, we have seen (§3.2) that when angular-momentum transport is viscous, then it is unlikely that stars can supply enough additional heat to stabilize the disk beyond one parsec, especially if viscosity is proportional to gas pressure rather than total pressure as viscous stability probably demands. In the latter case, even low-mass black holes embedded in the disk are probably inadequate. These statements assume that enough free gas remains in the disk to supply the central black hole at its Eddington rate. If all the gas converts to stars, stability may result, but the quasar is quenched. Perhaps a combination of stellar (or embedded-black-hole) heating and a super-viscous accretion speed may allow an extended gravitationally stable disk; we hope to explore this possibility in a future paper.

Given the serious theoretical difficulties of all proposed mechanisms for keeping $Q \gtrsim 1$ at large radii, we are forced to take seriously the only remaining possibility: that QSO disks do not exist much beyond $r_{\rm sg} \sim 10^{-2}\,$pc [eq. (9)]—at least not in a state of centrifugal support, vertical hydrostatic equilibrium, and steady accretion. Yet the mass within this radius is smaller than that of the central black hole by a factor $\sim (h/r) \sim 10^{-2}$ (since the midplane density $\approx M/2\pi r^3$ at $Q = 1$). Hence in order to grow the black hole, the disk must be replenished, either steadily or intermittently, by infall of low-angular-momentum material. In order that the gas not circularize outside $10^{-2} r_{\rm pc}$, its specific angular momentum must be $\lesssim 70\, M_8^{1/2}$ km s^{-1} pc, or some three orders of magnitude smaller than that of most stars in ellipticals and bulges (Binney & Merrifield 1998). Such a small ratio may seem unlikely, but on the other hand, $M_{\rm bh}/M_{\rm bulge} \approx 10^{-3}$ (McLure & Dunlop 2001). So perhaps the QSO is fueled by the low-angular-momentum tail of gas that forms the bulge. This gas would arrive at the outer edge of the disk in a vertically broad infall, perhaps already carrying dust formed from metals injected by outflows from the bulge or QSO disk itself, and hence taking the place of a warped outer disk as the source of reprocessed infrared light. The picture this calls to mind is similar, except in scale, to the standard scenario for the formation of a protostar (Shu et al. 1987).

I thank Charles Gammie, Pawan Kumar, Kristen Menou, Ramesh Narayan, and Bohdan Paczyński for helpful discussions. This work was supported by the NASA Origins Program under grant NAG 5-8385.

REFERENCES

Ayal S., Livio M., Piran T., 2000, ApJ, 545, 772

Balbus S. A., Hawley J. F., 1998, Rev. Mod. Phys., 70, 1

Balbus S. A., Papaloizou J. C. B., 1999, ApJ, 521, 650

Binney J., Merrifield M., 1998, Galactic astronomy. Princeton, NJ Princeton University Press

Blandford R. D., Payne D. G., 1982, MNRAS, 199, 883

Brandenburg A., Nordlund A., Stein R. F., Torkelsson U., 1995, ApJ, 446, 741+

Cameron A. G. W., 1978, in Protostars and Planets: Studies of Star Formation and of the Origin of the Solar System, pp 453–487

Cannizzo J. K., Lee H. M., Goodman J., 1990, ApJ, 351, 38

Cao X., Spruit H., 2001, astro-ph/0108484

Collin S., Zahn J., 1999a, A&A, 344, 433

Collin S., Zahn J., 1999b, Ap&SS, 265, 501

Ebisuzaki T., Makino J., Tsuru T. G., Funato Y., Portegies Zwart S., Hut P., McMillan S., Matsushita S., Matsumoto H., Kawabe R., 2001, ApJ, 562, L19

Evans C. R., Kochanek C. S., 1989, ApJ, 346, L13

Galeev A. A., Rosner R., Vaiana G. S., 1979, ApJ, 229, 318

Gammie C. F., 2001, ApJ, 553, 174

Gammie C. F., Narayan R., Blandford R., 1999, ApJ, 516, 177

Gebhardt K., Bender R., Bower G., Dressler A., Faber S. M., Filippenko A. V., Green R., Grillmair C., Ho L. C., Kormendy J., Lauer T. R., Magorrian J., Pinkney J., Richstone D., Tremaine S., 2000, ApJ, 539, L13

Greenhill L. J., Gwinn C. R., 1997, Ap&SS, 248, 261

Greenhill L. J., Jiang D. R., Moran J. M., Reid M. J., Lo K. Y., Claussen M. J., 1995, ApJ, 440, 619

Hawley J. F., 2000, ApJ, 528, 462

Heyvaerts J. F., Priest E. R., 1989, A&A, 216, 230

Hills J. G., 1975, Nature, 254, 295

Ichimaru S., 1977, ApJ, 214, 840

Illarionov A. F., Romanova M. M., 1988, AZh, 65, 682

Kalnajs A. J., 1971, ApJ, 166, 275+
Keady J. J., Kilcrease D. P., 2000, in Cox A. N., ed., , Allen's Astrophysical Quantities, 4 edn, AIP Press; Springer, New York, Chapt. 5, pp 95–120
Kumar P., 1999, ApJ, 519, 599
Kurucz R. L., 1992, Revista Mexicana de Astronomia y Astrofisica, 23, 181
Lacy J. H., Townes C. H., Hollenbach D. J., 1982, ApJ, 262, 120
Lee E., Goodman J., 1999, MNRAS, 308, 984
Lightman A. P., Eardley D. M., 1974, ApJ, 187, L1
Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1+
McLure R. J., Dunlop J. S., 2001, astro-ph/0108417
Maoz E., 1995, ApJ, 455, L131
Martin C. L., Kennicutt R. C., 2001, ApJ, 555, 301
Matsumoto R., Shibata K., 1997, in ASP Conf. Ser. 121: IAU Colloq. 163: Accretion Phenomena and Related Outflows, pp 443+
McMillan S. L. W., Lightman A. P., Cohn H., 1981, ApJ, 251, 436
Menou K., Quataert E., 2001, ApJ, 552, 204
Merloni A., Fabian A. C., 2001, MNRAS, 321, 549
Merritt D., Ferrarese L., 2001, ApJ, 547, 140
Miller K. A., Stone J. M., 2000, ApJ, 534, 398
Mukai K., Wood J. H., Naylor T., Schlegel E. M., Swank J. H., 1997, ApJ, 475, 812+
Nakai N., Inoue M., Miyoshi M., 1993, Nature, 361, 45
Narayan R., Yi I., 1994, ApJ, 428, L13
Neufeld D. A., Maloney P. R., 1995, ApJ, 447, L17
Ogilvie G. I., Livio M., 2001, ApJ, 553, 158
Paczyński B., 1978, Acta Astron., 28, 91
Ramsay G., Poole T., Mason K., Córdova F., Priedhorsky W., Breeveld A., Much R., Osborne J., Pandel D., Potter S., West J., Wheatley P., 2001, A&A, 365, L288
Rees M. J., 1978, The Observatory, 98, 210
Rees M. J., Phinney E. S., Begelman M. C., Blandford R. D., 1982, Nature, 295, 17
Sanders D. B., Phinney E. S., Neugebauer G., Soifer B. T., Matthews K., 1989, ApJ, 347, 29
Sellwood J. A., Balbus S. A., 1999, ApJ, 511, 660
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Shapiro S. L., Teukolsky S. A., 1985, ApJ, 292, L41
Shlosman I., Begelman M. C., 1987, Nature, 329, 810
Shlosman I., Begelman M. C., 1989, ApJ, 341, 685
Shu F. H., Adams F. C., Lizano S., 1987, ARAA, 25, 23
Spitzer L., 1978, Physical processes in the interstellar medium. New York : Wiley-Interscience
Spitzer L. J., Saslaw W. C., 1966, ApJ, 143, 400+
Stone J. M., Hawley J. F., Gammie C. F., Balbus S. A., 1996, ApJ, 463, 656+
Zel'Dovich Y. B., Podurets M. A., 1966, Soviet Astronomy, 9, 742+

APPENDIX A: VISCOUSLY HEATED DISKS

For completeness, this appendix gives formulae for the midplane temperature (T), the surface density (Σ), and the gravitational stability parameter (Q) in a steady disk heated by viscous dissipation only, and cooled by radiative diffusion.

Combining eqs. (1), (2), and (4), and writing $\beta c_{\rm s}^2 = k_{\rm B}T/m$, where $m \approx m_H$ is the mean mass per gas particle, we have the radial dependence of Σ & T:

$$T = \left(\frac{\kappa m}{16\pi^2\alpha\beta^{b-1}k_{\rm B}\sigma}\right)^{1/5} \dot{M}^{2/5}\Omega^{3/5} \qquad \text{(A1)}$$

$$\approx \quad 1.0\times 10^5\left(\frac{l_{\rm E}^2\hat\kappa}{\varepsilon_{0.1}^2\alpha_{0.01}\beta^{b-1}}\right)^{1/5} M_8^{-1/5}\left(\frac{10^3R_{\rm S}}{r}\right)^{9/10}\ {\rm K},$$

$$\Sigma = \frac{2^{4/5}}{3\pi^{3/5}}\left(\frac{m^4\sigma}{k_{\rm B}^4}\right)^{1/5}(\alpha\beta^{b-1})^{-4/5}\kappa^{-1/5}\dot M^{3/5}\Omega^{2/5} \tag{A2}$$

$$\approx \quad 3.9\times 10^6\,(\alpha_{0.01}\beta^{b-1})^{-4/5} l_{\rm E}^{3/5}\varepsilon_{0.1}^{-3/5}\hat\kappa^{-1/5}M_8^{1/5}\left(\frac{10^3R_{\rm S}}{r}\right)^{3/5}\ {\rm g\ cm^{-2}}\,.$$

If viscosity scales with gas pressure ($b=1$) then eqs. (A2)-(A1) do not depend on β, which in any case is not an independent parameter:

$$\frac{\beta}{1-\beta} = \frac{p_{\rm gas}}{p_{\rm rad}} = \frac{3ck_{\rm B}}{4\sigma m}\,\frac{\rho}{T^3} = \frac{3c}{8\sigma}\left(\frac{k_{\rm B}}{m}\right)^{1/2}\beta^{1/2}\frac{\Sigma\Omega}{T^{7/2}}\,;$$

this leads to

$$\frac{\beta^{(1/2)+(b-1)/10}}{1-\beta} \approx 0.44\,\alpha_{0.01}^{-1/10}\varepsilon^{4/5}l_{\rm E}^{-4/5}\left(\frac{r}{10^3R_{\rm S}}\right)^{21/20}. \tag{A3}$$

So the importance of gas pressure increases monotonically with radius in a steady, viscously heated disk.

Using eq. (A1) to eliminate $c_{\rm s}$ from eq. (6),

$$Q = \frac{3}{(4\pi)^{3/5}}\,\alpha^{7/10}\beta^{(7b-12)/10}\left(\frac{k_{\rm B}}{m\sigma^{1/4}}\right)^{6/5} G^{-1}\dot M^{-2/5}\Omega^{9/10} \tag{A4}$$

$$\approx \quad 8.1\times 10^{-2}\alpha_{0.01}^{7/10}\beta^{(7b-12)/10}\left(\frac{\varepsilon_{0.1}}{l_{\rm E}}\right)^{2/5}\hat\kappa^{3/10}M_8^{-13/10}\left(\frac{10^3R_{\rm S}}{r}\right)^{27/20}\,.$$